DOCUMENT RESUME

ED 078 074 TM 002 893

AUTHOR Browne, Michael W.
TITLE Gradient Methods for Analytic Rotation.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-22
PUB DATE Mar 73
NOTE 18p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Algorithms; Factor Analysis; Matrices; *Oblique Rotation; *Orthogonal Rotation; Statistical Analysis; Technical Reports

ABSTRACT

Gradient methods are employed in orthogonal oblique analytic rotation. Constraints are imposed on the elements of the transformation matrix by means of reparameterisations. (Author)

RESEARCH BULLETIN

RB-73-22

GRADIENT METHODS FOR ANALYTIC ROTATION

Michael W. Browne
Educational Testing Service
and
National Institute for Personnel Research
(South Africa)

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
March 1973

GRADIENT METHODS FOR ANALYTIC ROTATION

ABSTRACT

Gradient methods are employed in orthogonal and oblique analytic rotation. Constraints are imposed on the elements of the transformation matrix by means of reparameterisations.

GRADIENT METHODS FOR ANALYTIC ROTATION

1. INTRODUCTION

The analytic rotation of a factor matrix is a problem in optimisation subject to constraints. Given a $p \times m$ factor matrix, A, we have to find an $m \times m$ transformation matrix, T, which optimises a function, f, of the elements of the rotated factor matrix

$$\Lambda = AT \quad .$$

The transformation matrix is required to satisfy certain constraints;

$$T'T = I \tag{1}$$

in orthogonal rotation, and

$$\mathrm{Diag}(T^{-1}T^{-1'}) = I \tag{2}$$

in oblique rotation of the primary factor pattern.

If the reference structure rather than the primary factor pattern is to be rotated obliquely, other constraints are imposed on T. This approach, however, has serious disadvantages which have been pointed out by Jennrich & Sampson (1966). It will not be considered here.

Iterative algorithms for optimising a criterion for simple structure, f, which operate on pairs of columns of the factor matrix sequentially have been successful, both in orthogonal rotation (Kaiser, 1959) and in oblique rotation of the factor pattern (Jennrich & Sampson, 1966). Such algorithms do, however, have some disadvantages. Rounding errors can accumulate during iteration. Also, each step yields a conditional optimum of f with respect to one free parameter holding the rest fixed, a process which can sometimes converge slowly, particularly when the number of parameters is large (Box et al., 1969. p. 25).

Another algorithm for orthogonal rotation which obtains a new T on each iteration and does not accumulate rounding error has been proposed by Horst (1965, Sections 18.4, 18.7.2; Mulaik, 1972, Section 10.5). Little is yet known about convergence properties of the algorithm.

A gradient method with the property of quadratic termination, the Fletcher-Powell method, has been employed with considerable success by Jöreskog (1967, 1969) in maximum likelihood factor analysis. The purpose of this paper is to show that gradient methods, such as that of Fletcher & Powell (1963), can easily be employed in analytic rotation. Reparameterisations are used, the m^2 elements of T being expressed as functions of a smaller number of free parameters. Since a new T is obtained at each step of the algorithm, there is no accumulation of rounding error.

Section 2 reviews some basic results concerning the function to be minimised. A reparameterisation for orthogonal rotation and formulae for the gradient are given in Section 3. Corresponding results for oblique rotation are given in Section 4. In Section 5 the implementation of the procedures is discussed.

2. CRITERIA FOR SIMPLE STRUCTURE

The methods to be given subsequently are general and require only a criterion for simple structure to be minimised, f, and a corresponding $m \times m$ matrix of first derivatives

$$Z = \frac{\partial f}{\partial T} \quad .$$

A property of the criterion which will be assumed is that f is invariant under interchanges or reflections of columns of T.

We shall specifically consider a family of criteria for simple structure, dependent on a parameter κ ($0 \leq \kappa \leq 1$), which was proposed by Crawford & Ferguson (1970):

$$f = (1 - \kappa) \sum_{i=1}^{p} \sum_{j=1}^{m} \sum_{s \neq j}^{m} \lambda_{ij}^2 \lambda_{is}^2 + \kappa \sum_{j=1}^{m} \sum_{i=1}^{p} \sum_{r \neq i}^{p} \lambda_{ij}^2 \lambda_{rj}^2$$

$$= (1 - \kappa)\, \mathrm{tr}[D_\eta^2] + \kappa\, \mathrm{tr}[D_\gamma^2] - \mathrm{tr}[\Lambda'\Lambda^3] \qquad (3)$$

where

$$D_\eta = \mathrm{Diag}(\Lambda\Lambda') \quad ,$$

$$D_\gamma = \mathrm{Diag}(\Lambda'\Lambda) \quad ,$$

$$[\Lambda^3]_{ij} = \lambda_{ij}^3 \quad .$$

This family of criteria is equivalent to the Orthomax family in orthogonal rotation (Crawford & Ferguson, 1970, p. 324). Minimising f with $\kappa = 0$ gives the Quartimax solution, $\kappa = 1/p$ gives the Varimax solution, $\kappa = m/(2p)$ gives the Equamax solution and $\kappa = (m - 1)/(p + m - 2)$ gives the Parsimax solution.

In oblique rotation, minimisation of f (at least, when $0 \leq \kappa < 1$) cannot result in the factor correlation matrix,

$$C = T^{-1}T^{-1'} \quad ,$$

becoming singular (Crawford & Koopman, 1972). This result follows immediately from a theorem due to Jennrich & Sampson (1966, Theorem 1). Minimising f with $\kappa = 0$ gives the Quartimin solution considered by Jennrich & Sampson.

The $m \times m$ matrix of first derivatives of f with respect to the elements of T is (Mulaik, 1972, Section 10.5):

$$Z = \frac{\partial f}{\partial T} = 4A'\{(1 - \kappa)D_{\eta}\Lambda + \kappa\Lambda D_{\gamma} - \Lambda^3\} \quad . \tag{4}$$

It is of interest to note that f may be computed using

$$f = \frac{1}{4}\,\mathrm{tr}[T'Z] \tag{5}$$

which is equivalent to (3).

3. ORTHOGONAL ROTATION

In order to impose the $m(m + 1)/2$ constraints in (1) on the m^2 elements of T we shall express T as a function of

$$q = m^2 - m(m + 1)/2 = m(m - 1)/2 \tag{6}$$

parameters by means of the Cayley formulas (Gantmacher, 1959, pp. 288-289). If T is any orthogonal matrix such that

$$|I + T| \neq 0 \quad , \tag{7}$$

then there is a nonsingular matrix X, where

$$x_{ji} = -x_{ij} \quad , \qquad i > j \tag{8}$$

$$x_{ii} = 1 \quad , \qquad i = 1,2 \ldots m \tag{9}$$

such that

$$T = 2X^{-1} - I \quad . \tag{10}$$

Conversely, given any nonsingular X satisfying (8) and (9), the matrix T constructed from (10) will satisfy (1).

We can therefore regard the transformation matrix T as a function of a matrix X with elements satisfying (8) and (9) and minimise f with respect to the $m(m - 1)/2$ free parameters $x_{21}, x_{31}, x_{32} \cdots x_{m,m-1}$.

If necessary, the matrix X corresponding to a particular T may be obtained from

$$X = 2(I + T)^{-1} \quad . \tag{11}$$

Because of (7), orthogonal matrices such as

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

which result in interchanges or reflections of columns of A cannot be represented by (10). This, however, does not matter since f is invariant under such operations.

Gradient methods for minimising f will require first derivatives of f with respect to the x_{ij} $(i < j)$. Using the chain rule we obtain:

$$\frac{\partial f}{\partial x_{ij}} = \mathrm{tr}\left[\frac{\partial f}{\partial T'} \frac{\partial T}{\partial x_{ij}} \right]$$

$$= -2\mathrm{tr}[Z'X^{-1} \frac{\partial X}{\partial x_{ij}} X^{-1}]$$

$$= 2([X^{-1}Z'X^{-1}]_{ij} - [X^{-1}Z'X^{-1}]_{ji}) \quad . \tag{12}$$

This result is general. Simplification is possible when Z is defined by (4). It is easily verified that $X^{-1}\Lambda'D_\eta AX^{-1}$ is symmetric. Consequently the first term in (4) may be discarded and (12) becomes

$$\frac{\partial f}{\partial x_{ij}} = 2([X^{-1}WX^{-1}]_{ij} - [X^{-1}WX^{-1}]_{ji}) \tag{13}$$

where

$$W = 4\{\kappa D_\gamma \Lambda' - (\Lambda^3)'\}A \quad .$$

This simplification reflects the fact that the first term in (3) remains constant when T is orthogonal.

4. OBLIQUE ROTATION

The m constraints in (2) may be imposed on the m^2 elements of T by expressing T as a function of

$$q = m^2 - m = m(m - 1) \tag{14}$$

parameters.

If T is any nonsingular matrix which satisfies (2) and which has nonnegative diagonal elements,

$$t_{ii} > 0 \quad , \qquad i = 1,2, \ldots m \tag{15}$$

then there is a nonsingular matrix X, with diagonal elements satisfying (9), such that

$$T = X \,\mathrm{Diag}^{\frac{1}{2}}(V) \tag{16}$$

where

$$V = X^{-1} X^{-1'} \quad .$$

Conversely, given any nonsingular X, the matrix T constructed using (16) will satisfy (2).

We may therefore define the transformation matrix, T, by (16) and minimise f with respect to the $m(m - 1)$ nondiagonal elements of X. Again, certain permutations or reflections of columns of T cannot be represented because of (15). Given any <u>nonsingular</u> T, however, it is always possible to interchange and reflect columns so that (15) is satisfied.

The factor correlation matrix is

$$C = \mathrm{Diag}^{-\frac{1}{2}}(V)\; V \;\mathrm{Diag}^{-\frac{1}{2}}(V)$$

and, if necessary, X may be obtained from T using

$$X = T \,\mathrm{Diag}^{-1}(T) \quad . \tag{17}$$

The gradient is given by

$$\frac{\partial f}{\partial x_{ij}} = \mathrm{tr}\left[\frac{\partial f}{\partial T'}\frac{\partial T}{\partial x_{ij}}\right] \qquad (i \neq j)$$

$$= \mathrm{tr}\left[Z'\left\{\frac{\partial X}{\partial x_{ij}}\,\mathrm{Diag}^{\frac{1}{2}}(V) - X\,\mathrm{Diag}^{-\frac{1}{2}}(V)\,\mathrm{Diag}\left(X^{-1}\frac{\partial X}{\partial x_{ij}}V\right)\right\}\right]$$

$$= [Z\,\mathrm{Diag}^{\frac{1}{2}}(V) - X^{-1'}\,\mathrm{Diag}(Z'X)\,\mathrm{Diag}^{-\frac{1}{2}}(V)\,V]_{ij} \quad . \qquad (18)$$

5. EXPERIENCE WITH GRADIENT METHODS

In employing gradient methods, the criterion for simple structure, f, is regarded as a function of the q free elements of the matrix X, where q is defined by (6) or (14). Given X, the transformation matrix T is obtained from (10) or (16), the elements, $\partial f/\partial x_{ij}$, of the gradient are obtained from (13) or (18) and (4), and f is obtained from (3) or from (5).

Two gradient methods were tried. The first was Fletcher's (1970) method, a development of the Fletcher-Powell (1963) method which seems to require fewer function evaluations. The second was the Polak-Ribière method (Polak, 1971) which is similar to the conjugate gradient method of Fletcher & Reeves (1964). Fletcher's method, like that of Fletcher & Powell, builds up an inverse Hessian matrix. The Fletcher-Reeves and Polak-Ribière methods do not. Consequently, they require less computer storage than the Fletcher and Fletcher-Powell methods but appear to converge less rapidly. When the number of factors, m, is less than six or seven, there are advantages in employing an algorithm which builds up an inverse Hessian like that of Fletcher; when m is large, storage considerations could require the use of the Fletcher-Reeves or Polak-Ribière methods.

The Fletcher method was implemented by making minor changes to a subroutine package prepared by Gruvaeus & Jöreskog (1970) for the Fletcher-Powell method. As suggested by Gruvaeus & Jöreskog (1970), a starting point for the Fletcher method was obtained by carrying out a few initial steepest descent iterations, starting with $X = I$. The steepest descent iterations were terminated when two consecutive iterations yielded a relative decrease in f of less than five percent.

In applications the algorithm has appeared to be satisfactory. Table 1 shows the orthogonal and oblique factor matrices obtained for Harman's 24

Insert Table 1 about here

---

psychological tests (Harman, 1960, Table 10.10) with $\kappa = 1/p$ (Varimax). Row normalisation of A (e.g., Harman, 1960, p. 302) was not carried out. Table 2 gives the primary factor correlation matrix C for the oblique solution. Details of the iterations are shown in Table 3. It can

---

Insert Tables 2 and 3 about here

---

be seen that the oblique rotation required more iterations than the orthogonal rotation. This was found to be true in general and can be expected since more free parameters are involved in oblique rotation. Iteration was terminated when all elements of the gradient vector were less than $.00001 \times 24$ in absolute value.

In implementing the Polak-Ribière method (Polak, 1971, pp. 53-54) the linear search subroutine of Gruvaeus & Jöreskog (1970) was employed. Reinitialisation of the process with a steepest descent step was carried out after each set of $q + 1$ iterations. Inequalities (7) and (15) suggest that one should ensure, before each reinitialisation, that the ordering of columns T maximises $\Pi_{i=1}^{p} |t_{ii}|$ and that $t_{ii} > 0$, $i = 1 \ldots m$. This was done and, if reordering or reflection of columns of T was necessary, X was recomputed using (11) or (17). In oblique rotation in particular this step improved convergence. The Polak-Ribière method then appeared to be not much slower than the Fletcher method.

Using the same starting point and convergence criterion as those of the Fletcher method, the Polak-Ribière method was applied to Harman's factor matrix. The rotated factor matrices yielded by the two methods agreed to three decimal places. It can be seen from Table 3 that the Polak-Ribière method compared quite favourably with the Fletcher method in speed of convergence.

ACKNOWLEDGMENTS

My thanks go to Ms. M. van Thillo who checked my algebra and prepared all computer programs for the procedures described here.

I am indebted, also, to Mr. G. T. Gruvaeus, to Dr. T. Stroud and to Professor L. R Tucker for valuable suggestions which have been implemented in this work.

REFERENCES

BOX, M. J., DAVIES, D. & SWANN, W. H. (1969). Non-linear Optimisation Techniques. Edinburgh: Oliver & Boyd.

CRAWFORD, C. B. & FERGUSON, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. Psychometrika 35, 321-332.

CRAWFORD, C. B. & KOOPMAN, P. (1972). A general method of rotation and its application in oblique primary factor rotation. Unpublished paper. Vancouver, B.C.: Simon Fraser University.

FLETCHER, R. (1970). A new approach to variable metric algorithms. Comp. J. 13, 317-322.

FLETCHER, R. & POWELL, M. J. D. (1963). A rapidly convergent descent method for minimization. Comp. J. 6, 163-168.

FLETCHER, R. & REEVES, C. M. (1964). Function minimisation by conjugate gradients. Comp. J. 7, 149-154.

GANTMACHER, F. R. (1959). The Theory of Matrices, vol. 1. New York: Chelsea.

GRUVAEUS, G. T. & JÖRESKOG, K. G. (1970). A computer program for minimizing a function of several variables. Research Bulletin 70-14. Princeton, N.J.: Educational Testing Service.

HARMAN, H. H. (1960). Modern Factor Analysis. Chicago: University of Chicago Press.

HORST, P. (1965). Factor Analysis of Data Matrices. New York: Holt, Rinehart & Winston.

JENNRICH, R. I. & SAMPSON, P. (1966). Rotation for simple loadings. Psychometrika 31, 313-323.

JÖRESKOG, K. G. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika 32, 443-482.

JÖRESKOG, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika 34, 183-202.

KAISER, H. F. (1959). Computer program for varimax rotation in factor analysis. Educ. psychol. Measmt 19, 413-420.

MULAIK, S. A. (1972). The Foundations of Factor Analysis. New York: McGraw-Hill.

POLAK, E. (1971). Computational Methods in Optimization. New York: Academic Press.

Table 1.—Harman's 24 Psychological Tests:

Factor Pattern Matrices. $\kappa = 1/p$.

Orthogonal Rotation				Oblique Rotation			
.235	.653	.197	.125	.002	.657	.102	.088
.165	.419	.072	.074	.024	.429	.007	.054
.223	.520	.020	.054	.068	.545	-.058	.028
.267	.518	.086	.037	.110	.535	.014	.001
.780	.124	.208	.052	.750	.045	.141	-.003
.785	.135	.098	.141	.746	.056	.001	.107
.843	.105	.153	-.002	.841	.036	.092	-.062
.594	.309	.254	.048	.493	.258	.188	-.009
.837	.120	.013	.184	.807	.039	-.103	.161
.170	-.075	.714	.166	.075	-.212	.723	.124
.218	.065	.627	.290	.069	-.071	.583	.264
.062	.228	.692	.040	-.099	.154	.708	-.024
.240	.390	.590	-.014	.065	.341	.582	-.092
.261	.017	.196	.463	.146	-.103	.080	.496
.174	.128	.107	.478	.029	.035	-.023	.520
.171	.405	.127	.408	-.044	.349	-.012	.428
.197	.053	.232	.607	.032	-.086	.086	.658
.082	.323	.306	.506	-.157	.227	.172	.532
.189	.224	.180	.361	.032	.150	.072	.377
.423	.432	.116	.205	.260	.396	.003	.185
.229	.402	.394	.202	.029	.343	.319	.170
.435	.371	.064	.315	.274	.316	-.074	.316
.432	.526	.214	.159	.239	.495	.109	.121
.398	.182	.459	.267	.250	.070	.383	.238

Table 2.—Harman's 24 Psychological Tests:

Primary Factor Correlations. $\kappa = 1/p$.

1.000			
.381	1.000		
.271	.295	1.000	
.350	.330	.375	1.000

Table 3.—Details of the Iterations

	Orthogonal Rotation	Oblique Rotation
No. of initial steepest descent iterations. Fletcher & Polak-Ribière.	4	2
No. of iterations. Fletcher.	9	31
No. of function evaluations. Fletcher.	15	54
No. of iterations. Polak-Ribière.	10	39
No. of function evaluations. Polak-Ribière.	19	75